

B21  
PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : <b>C12Q 1/68</b>		A1	(11) International Publication Number: <b>WO 00/50640</b> (43) International Publication Date: <b>31 August 2000 (31.08.00)</b>
(21) International Application Number: <b>PCT/US00/01759</b> (22) International Filing Date: <b>24 January 2000 (24.01.00)</b>		(81) Designated States: AU, CA, JP, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).	
(30) Priority Data: 60/122,177 25 February 1999 (25.02.99) US		Published <i>With international search report. Before the expiration of the time limit for attending the claims and to be republished in the event of the receipt of amendments.</i>	
(71) Applicant: EXACT LABORATORIES, INC. (US/US); 63 Great Road, Maynard, MA 01754 (US).			
(72) Inventors: SHUBER, Anthony, P.; 11 Grant Street, Milford, MA 01757 (US). HUNTRASS, Frederick, A., Jr.; 396 Wilder Road, Bolton, MA 01740 (US). MOORE, James, K.; 13 Conant Drive, Stow, MA 01775 (US).			
(74) Agent: MEYERS, Thomas, C.; Testa, Hurwitz & Thibeault, LLP, High Street Tower, 125 High Street, Boston, MA 02110 (US).			
(54) Title: METHODS FOR PRESERVING DNA INTEGRITY			
(57) Abstract <p>Methods for extracting DNA from a biological sample that result in a higher yield of target DNA than conventional methods. More particularly, methods for extracting DNA include exposing the biological sample to inhibitors of DNA degradation.</p>			

***FOR THE PURPOSES OF INFORMATION ONLY***

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Larvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

## Methods for Preserving DNA Integrity

### Cross-Reference to Related Application

The present application claims priority to and the benefit of U.S. provisional patent application serial number 60/122,177, filed February 25, 1999, the entire disclosure of which is incorporated herein by reference.

### Technical Field

5        The invention provides methods for deoxyribonucleic acid ("DNA") extraction from a biological sample. More particularly, the invention relates to methods for high yield DNA extraction from a heterogeneous biological sample by inhibiting DNA degradation.

### Background of the Invention

DNA is a relatively stable molecule that is routinely isolated from biological samples.

10      Recently, many diseases involving instabilities (e.g., mutations) in genomic DNA have been characterized. Also, many pathogens have been identified by the presence or absence of a particular DNA in a biological sample. Many diseases, such as cancer, are optimally detected early in their progression. In order for early detection to be effective, relatively low levels of DNA which are indicative of cancer must be detected against a high background of other DNA

15      (e.g., normal human DNA, bacterial DNA, etc.). This type of detection is technically difficult and typically results in low sensitivity of detection. Moreover, in certain complex specimens, including stool, what little species-specific DNA exists, is rapidly degraded, making efficient sequence-specific detection even more difficult. Thus, a need exists for methods to retain integrity of DNA in a sample, especially in samples in which the DNA to be detected is in low

20      proportion relative to other DNA in the sample and is degraded quickly.

Summary of the Invention

The present invention provides methods for preserving the integrity of DNA in a sample.

In a preferred embodiment, methods of the invention prevent enzyme-mediated DNA degradation. Preservation of DNA integrity facilitates isolation and detection of DNA.

Methods of the invention are especially useful for extracting or detecting DNA in a biological specimen, especially one that contains low levels of relevant DNA. A good example of a specimen that contains lower-levels of relevant DNA is stool. Typical human stool contains only small amounts of intact human DNA. Most of the human DNA in stool from a healthy individual is presumably from exfoliated epithelial cells, and has undergone apoptotic degradation. As the forming stool passes through the colon, colonic epithelial cells are sloughed onto the stool as part of the cellular turnover that occurs in the colon. Stool also contains sloughed cells from other luminal sources (e.g., lung, stomach, esophagus, etc.) Sloughed cells typically have undergone or are undergoing apoptosis, leaving cellular DNA in small fragments. Enzymes, such as deoxyribonuclease ("DNase") and Micrococcal nuclease contribute to the degradation of any intact human DNA that remains. Prior art methods, while using DNase inhibitors, have failed to achieve significant yields of intact, species-specific DNA from stool. Therefore, such methods failed to consider optimization of inhibition of DNA degradation. Methods of the invention are based on the realization that optimal inhibition of DNA degrading enzyme(s) effectively preserves DNA, especially large, diagnostically-relevant DNA fragments that are present in a sample.

In one aspect, the invention comprises inhibiting nucleic acid degradation in a sample and optionally extracting a target DNA with, for example, a phenol-chloroform extraction. Preferably, the inhibition of nucleic acid degradation is sufficient to produce a critical number of

molecules of analyzable DNA. In one embodiment, methods of the invention comprise inhibiting an enzyme capable of DNA degradation in a stool sample. In a preferred embodiment, methods of the invention comprise exposing a stool sample to an ion chelator, such as a divalent ion chelator. Ion chelators, in certain embodiments inhibit DNase. Examples of preferred 5 inhibitors include ethylenediaminetetraacetic acid ("EDTA"). Additional preferred methods of the invention comprise exposing a stool sample to a Micrococcal nuclease inhibitor, such as EGTA, also a divalent ion chelator. Inhibitors of DNA degradation may be used either alone or in combination to achieve optimal levels of DNA preservation.

Methods of the invention are practiced using any inhibitor of DNA degradation. The 10 amount of inhibitor varies depending on the inhibitor that is used. However, an inhibitor must be used in an amount that preserves significant levels of DNA in the sample for subsequent analysis. Methods for determining sufficient levels of DNA are presented below. Such methods allow the skilled artisan to practice the invention with specificity regardless of the inhibitor used. According to preferred methods, an amount of inhibitor is used that preserves sufficient DNA in 15 the sample for detection of a target DNA within a desired level of statistical confidence. Using methods described herein, the skilled artisan can determine an appropriate amount of any inhibitor for use in methods of the invention. The use of various specific inhibitors is exemplified below.

In another preferred embodiment, methods of the invention comprise obtaining a 20 representative (circumferential or cross-sectional) stool sample, exposing the sample or a portion thereof to a DNase inhibitor, and isolating DNA from the sample. One preferred DNase inhibitor is EDTA. Preferred amounts of EDTA are from about 0.042 g per gram of stool to about 0.782 g per gram of stool and especially from about 0.250 g per gram of stool to about

0.521 g per gram of stool. DNA may be extracted, for example, by a phenol-chloroform extraction. After extraction, the DNA may be analyzed by methods known in the art. For example, U.S. Patent No. 5,830,665 and U.S. Patent No. 5,670,325, which are incorporated by reference herein, disclose methods for analyzing DNA which has been extracted from a stool sample.

Methods of the invention are useful in any sample in which inhibition of DNA degradation is desired. For example, methods of the invention are especially effective in samples comprising exfoliated cells, especially exfoliated epithelial cells. The DNA contained in such samples typically degrades rapidly, making analysis of a particular DNA, especially one that exists in low proportion within the sample, difficult. For example, such samples include stool, sputum, urine, pus, and colostrum. Methods of the invention include inhibiting DNA degradation in such samples, thus preserving a sufficient amount of DNA for specific, sensitive detection. Any of the features described above, such as DNA degradation inhibitors or amounts of inhibitors that are used, can be useful in samples containing exfoliated cells.

**15 Brief Description of the Drawings**

FIG. 1 shows a flow chart describing one aspect of the invention.

FIG. 2 shows a separation gel of DNA isolated from several different homogenized stool supernatants that contained various concentrations of EDTA.

FIG. 3 shows a separation gel of DNA isolated from homogenized stool supernatant that contained various concentrations of EDTA followed by capture and amplification.

- 5 -

FIG. 4 shows separation gel of DNA isolated from homogenized stool supernatant that contained various concentrations EDTA followed by capture, addition of more DNA, and amplification.

FIG. 5 shows a set of curves produced by regression analysis of the data obtained using  
5 the model as described for, for example, Tables 1 and 2.

#### Detailed Description of the Invention

##### I. Introduction

The present invention provides methods for increased yield of DNA in a biological sample by preserving the integrity of DNA in the sample. Such methods are especially useful  
10 when the DNA of interest ("the target DNA") is present in the sample at a low frequency, or is rapidly degraded. More particularly, methods of the invention include, for example, inhibiting enzymes that degrade DNA.

Prior to the present invention, those skilled in the art have not been concerned with preventing DNA degradation prior to extraction from a sample. Typically, either the DNA of  
15 interest is present in samples in relatively large quantities (e.g., tumor cells, blood), or methods are directed toward increasing sensitivity to low-frequency DNA, and not to preserving its integrity. However, especially in the case of low-frequency DNA in a heterogeneous sample (e.g., a sample having cells and/or cellular debris from multiple cell types and/or organisms), methods for increasing sensitivity to DNA have not been entirely successful. Methods of the  
20 invention provide a new approach by preserving the integrity of DNA. Methods of the invention increase the likelihood of detecting a specific target DNA, because such methods make more intact target DNA available in the sample.

Referring to Figure 1, one generalized method of the invention involves obtaining a sample (step 2) and exposing it to a DNA degradation inhibitor (step 4). Once the sample has been exposed to the DNA degradation inhibitor, target DNA is extracted (step 6). The presence or absence of this extracted target DNA is then detected (step 8).

5 In heterogeneous samples, such as stool, endogenous human DNases and/or bacterial nucleases degrade DNA. Examples of nucleases include DNases, such as deoxyribonuclease I ("DNase I") and Micrococcal nuclease. DNase and Micrococcal nuclease both require a divalent cation to function optimally. For DNase, suitable ions include  $Mn^{+2}$  and  $Mg^{+2}$ . For Micrococcal nuclease,  $Ca^{+2}$  is a suitable ion. Ion chelators, and particularly divalent ion chelators, are capable 10 of inhibiting nucleases. Ion chelators remove ions from association with the nuclease, thus inhibiting the nuclease's function. For example, EDTA or EGTA used in optimal amounts are useful ion chelators for use in the present invention.

Other compounds that inactivate, interfere with, or slow enzyme-mediated degradation of 15 DNA are useful. For example, ligands and/or antibodies which compete for or interfere with the active site of DNA degrading enzymes, which inactivate those enzymes, and/or which block messenger systems that control DNA degrading enzymes are useful in the practice of the invention. Phenol-chloroform extraction components, at higher concentrations than those typically used during extraction, also are capable of inhibiting nucleases by separation and 20 denaturation. For example, phenol denatures DNA degradation enzymes, and is used in methods of the invention to preserve DNA integrity. Also, proteinases which degrade and/or denature DNA degrading enzymes are useful.

Methods of the invention comprise the use of optimal amounts of DNA degradation inhibitors in order to preserve high-integrity DNA sufficient for diagnostic screening. In

heterogeneous samples, such as stool, target DNA (e.g., mutated DNA or reductions in enumerated wild type DNA that are indicative of a mutation) is present in low amounts. An optimal amount of a DNA degradation inhibitor is an amount that results in a measurable improvement in the quantity of DNA available in the sample. Thus, the skilled artisan can 5 empirically determine optimal amounts of DNA degradation inhibitors for use in methods of the invention by using inhibitor amounts necessary to preserve a diagnostically-relevant fraction of high-integrity target DNA. A method for determining diagnostically-relevant DNA amounts is presented below.

Amplification of DNA, and other stochastic processes, performed on heterogeneous 10 samples may actually contribute to the inability to measure low-frequency DNA. For example, a typical cancer-associated (mutant) DNA in the early stages of oncogenesis represents about 1% of the DNA in a heterogeneous sample (e.g., stool). If DNA in the sample is amplified at 30% PCR efficiency, any particular DNA has only a 30% chance of being amplified in any round of PCR. Thus, if a mutant DNA initially present as 1% of a sample is not amplified in the first 15 round, the mutant DNA will represent only about 0.7% of the DNA in the sample after round 1. If no mutant is amplified in the first two rounds ( $0.7 \times 0.7$ , or a 49% probability), the mutant DNA will represent only about 0.6% of the DNA in the sample going into round three of the PCR. If the post-amplification assay used to detect the mutant has a sensitivity of no more than 0.5% for the mutant, it may not be possible to reliably detect the presence of the mutant DNA. 20 Thus, the detection method itself may actually contribute to difficulties in detecting low-frequency DNA, especially if sufficient amounts of intact DNA are not present in a sample. Thus, one means for determining an appropriate amount of inhibitor to use in methods of the invention is to determine the minimum amount of intact DNA that must be present in a sample to

avoid the stochastic effects described above, and then to use sufficient inhibitor to produce at least the minimum number of DNA molecules in the sample. Methods for calculating the minimum number of DNA molecules necessary to overcome the effects of stochastic processes, such as PCR, are presented below.

5        A model useful to generate sufficient DNA molecules for accurate measurement operates by iterating stochastic processes over a number of rounds of PCR. In the context of molecular disease diagnostics, the model dictates the number of molecules that must be presented to the PCR in order to reliably ensure amplification of desired target DNA. The model incorporates a preset PCR efficiency (established to meet separate specificity requirements), and a preset ratio  
10      of mutant DNA to total DNA in the sample to be analyzed (which is a property of the disease to be detected and the nature of the sample). Based upon those input values, the model predicts the number of molecules that must be presented to the PCR in order to ensure, within a predefined level of statistical confidence, that a low-frequency (target) molecule will be amplified and detected. Once the number of molecules is determined, the skilled artisan can determine the  
15      sample size to be used (e.g., the weight, volume, etc.), depending on the characteristics of the sample (e.g., its source, molecular makeup, etc.). The model dictates the number of molecules that must be presented to the PCR in order to reliably ensure amplification and detection.

      The exemplary model simulates selection of DNA for amplification through several rounds of PCR. For purposes of the model, a sample is chosen that contains a ratio of mutant-to-  
20      total DNA of 1:100, which is assumed to lie at the clinical threshold for disease. For example, in colorectal cancer 1% of the human DNA in a specimen (e.g., stool) is mutated (i.e., has a deletion, substitution, rearrangement, inversion, or other sequence that is different than a corresponding wild-type sequence). Over a large number of PCR rounds, both the mutant and

wild-type molecules will be selected (*i.e.*, amplified) according to their ratio in the specimen (here, nominally 1 in 100), assuming there are any abnormal molecules in the sample. However, in any one round, the number of each species that is amplified is determined according to a Poisson distribution. Over many rounds, the process is subject to stochastic errors that reduce 5 the ability to detect low-frequency mutant DNA. However, the earlier rounds of PCR (principally, the first two rounds) are proportionately more important when a low-frequency species is to be detected, and any rounds after round 10 are virtually unimportant. Thus, the model determines the combined probability of (1) sufficient mutant molecules being presented to the PCR, and (2) the effects of stochastic amplification on those molecules so that at the output 10 of the PCR there will be a sufficient number of molecules and a sufficient ratio of mutant to total molecules to assure reliable detection.

The model used to run the number of molecules necessary at the first round of PCR was generated as a "Monte Carlo" simulation of a thousand experiments, each experiment consisting of 10 cycles of PCR operating on each molecule in the sample. The simulation analyzed (1) 15 taking a sample from the specimen; and (2) each round of PCR iteratively to determine whether, for each round, a mutant DNA if present in the sample was amplified. Upon completion of the iterative sampling, the model determined the percent of rounds in which a mutant strand was amplified, the percent of mutants exceeding a predetermined threshold for detection (in this example 0.5% based upon the mutant:total ratio of 1%), the coefficient of variation (CV) for 20 stochastic sampling in each round alone, and the coefficient of variance for stochastic sampling and PCR in combination.

Stochastic noise is created in PCR if the PCR efficiency is anything other than 0% or 100% (these two cases represent either there is no amplification at all or perfect fidelity of

specific amplification). The noise, or background, signal level in a PCR that is between 0% and 100% varies with the efficiency of the PCR. The standard deviation of stochastic noise, S, in a PCR is given by the equation,  $S = \sqrt{npq}$ , where n is the number of molecules in the sample, p is the efficiency of PCR, and q is 1-p. Table 1 presents results obtained for iterative samplings 5 with PCR efficiency set at 100% and 20%, and a mutant:total ratio of 0.5%.

Table 1 represents output from the model in 12 experiments conducted under various conditions. The first row shows the nominal number of molecules entering the first round of PCR (i.e., the total number of molecules available for amplification). The second row shows the percent of molecules (DNA) in the biological specimen that is expected to be mutant. For 10 colorectal cancer indicia in DNA recovered from stool, the threshold for clinical relevance in the detection of early stage cancer is 1%. That is, 1% of the DNA in a sample derived from a heterogeneous specimen (e.g., stool) contains a mutation associated with colorectal cancer. The 6th row is the threshold of detection of the assay used to measure PCR product after completion of PCR. That number is significant, as will be seen below, because sufficient mutant DNA must 15 be produced by PCR to be detectable over aberrant signal from wild-type and random background noise. Under the heading "Outputs", the first line provides the likelihood that at least one mutant molecule is presented to the first round of PCR. The second line under the Output heading provides the likelihood of detection of mutants (after PCR) above the predetermined threshold for detection. For example, in experiment 4, the results indicate that in 20 87.9% of experiments run under the conditions specified for experiment 4, the number of mutants will exceed the threshold number for detection. Finally, the last two rows provide the coefficient of variation for sampling, and for the combination of sampling and PCR.

TABLE 1

Inputs	100% Efficiency PCR						20% Efficiency PCR					
	Exp 1	Exp 2	Exp 3	Exp 4	Exp 5	Exp 6	Exp 7	Exp 8	Exp 9	Exp 10	Exp 11	Exp 12
Nominal number of Molecules going into PCR	50	100	200	500	1,000	10,000	50	100	200	500	1,000	10,000
Percent of molecules that are mutant	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%	1%
Number of PCR Rounds	10	10	10	10	10	10	10	10	10	10	10	10
Efficiency of PCR per round	100% + 100%	100% + 100%	100% + 100%	100% + 100%	100% + 100%	100% + 100%	20% + 20%	20% + 20%	20% + 20%	20% + 20%	20% + 20%	20% + 20%
Number of experiments modeled	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
Min. Molecules which can be reliably detected by assay	0.50%	0.50%	0.50%	0.50%	0.50%	0.50%	0.50%	0.50%	0.50%	0.50%	0.50%	0.50%
Outputs												
% of Experiments each with the number of mutant cells > 0	39.30%	64.40%	87.30%	99.40%	100%	100%	39.10%	63.50%	86.10%	99.60%	100%	100%
% of Experiments each with number of mutant cells >= 200 (detection threshold)	39.30%	64.40%	87.30%	97.90%	100%	100%	38.20%	52.40%	63.30%	81.60%	100%	100%
Stochastic Sampling CV	143.20% + 100%	70%	44.10% + 32.10%	9.9%	144.80% + 101.80%	72.00%	44.20%	32.00%	10%	179.70%	123.10%	82.10%
Stochastic Sampling and Stochastic PCR CV	143.20% + 100%	70%	44.10% + 32.10%	9.9%	144.80% + 101.80%	72.00%	44.40%	40.30%	12.70%	179.70%	123.10%	82.10%

5 As shown in Table 1, even at 100% PCR efficiency, mutant DNA is detected in only 97.1% of the samples when 1000 input molecules are used (*i.e.*, 1000 DNA molecules are available for priming at the initial PCR cycle), even though 100% of the DNA is amplified in any given round of PCR. When 10,000 molecules are presented, it is virtually certain that the mutant DNA will be amplified and detected, as shown in the results for experiment 6 in Table 1.

10 Stochastic errors due to variation in the number of input molecules become less significant at about 500 input molecules and higher (*i.e.*, the CV for stochastic variations is about the same regardless of whether PCR efficiency is 20% or 100%). At lower PCR efficiency (20% in Table 1), the model shows that introducing 50, 100, 200, 500, or even 1000 molecules into the PCR does not assure either amplification or detection. As shown in experiment 12, introducing

15 10,000 molecules results in amplification of the mutant target, and a high likelihood of its subsequent detection. Thus, even with 100% efficient PCR, significant false negative events occur when input molecules fall below 500.

The foregoing analysis shows that there is a unique range for the number of molecules that must be presented to a PCR in order to achieve amplification of a low-frequency DNA, and to allow its detection. That range is a function of the PCR efficiency, and the percentage of low-frequency (mutant) DNA in the sample, and the detection threshold. The aforementioned model 5 was developed and run in Visual Basic for Applications code (Microsoft, Office 97) to simulate a PCR as described above. The statistical confidence level within which results were measured was held constant at approximately 99%. Only the PCR efficiency and percent mutant DNA were varied. As discussed above, the model iteratively samples DNA in a "Monte Carlo" simulation over a thousand experiments, each experiment consisting of 10 rounds of PCR. The 10 results are shown below in Table 2.

TABLE 2

PCR Efficiency	Number of molecules needed			
	1% Mutant	2% Mutant	5% Mutant	10% Mutant
10%	3,000			
20%	2,500			
50%	2,200			
100%	1,600			
		1,500		
10%		1,200		
20%		1,000		
50%		800		
			500	
10%			450	
20%			400	
50%			300	
				225
10%				200
20%				150
50%				125
100%				

Regression of the data obtained using the model as described above produced the set of curves set forth below in Figure 5.

Using Figure 5, the optimal number of molecules to be presented to the PCR is determined by selecting a PCR efficiency (or determining the efficiency by empirical means), and selecting a percentage of the sample suspected to be mutant DNA associated with disease. This, in turn, dictates a threshold of detection. Not all detection strategies have similar underlying detection thresholds, so an appropriate technology must be selected. The percentage mutant DNA may be determined by clinical considerations as outlined above for colorectal cancer.

One may determine the PCR efficiency and percent expected mutant in order to maximize the probability of obtaining amplified, detectable mutant DNA. For example, one may select N, the number of input molecules from the "1%" curve in Figure 5, when 5% of the

sample is expected to be mutant DNA in order to increase the confidence of the assay result.

This model, and particularly Figure 5, are useful when determining optimal concentrations of a DNA degradation inhibitor. If PCR is used to analyze DNA after the DNA sample is exposed to a DNA inhibitor, Figure 5, will indicate how many molecules of target 5 DNA need to be preserved in order to have sufficient analyzable DNA. Thus, the optimal amount of any DNA degradation inhibitor may be determined as that amount, or range of amounts, of inhibitor that produce a sufficient number of analyzable DNA according to Figure 5. Of course, this modeling system may be applied to DNA detection techniques other than PCR. Specifically, those skilled in the art can apply this modeling system to any process and/or 10 detection technique in which stochastic noise is problematic. Thus, the optimal amount of any DNA degradation inhibitor can be determined based upon the number of DNA molecules that are sufficient to produce analyzable DNA.

Once the number of molecules for input to the PCR is determined, a sample comprising that number of molecules (or greater) is prepared for PCR according to standard methods. The 15 number of molecules in a sample may be determined directly by, for example, enumerative methods such as those taught in U.S. Patent No. 5,670,325, incorporated by reference herein. Alternatively, the number of molecules in a complex sample may be determined by molar concentration, molecular weight, or by other means known in the art. The amount of DNA in a sample may be determined by mass spectrometry, optical density, or other means known in the 20 art. The number of molecules in a sample derived from a biological specimen may be determined by numerous means in the art, including those disclosed in U.S. Patent Nos. 5,741,650 and 5,670,325, both of which are incorporated by reference herein.

Methods as described above are used to determine minimum or optimal amounts of DNA

degradation inhibitors for use in any DNA isolation, detection, or amplification process in which stochastic processes occur. Using the above-described model for determining the minimum number of molecules that must be measured to reliably detect a low-frequency species, one can empirically determine how much of any given inhibitor should be used.

5 II. Example - Detection of DNA in Stool with EDTA as a DNA Degradation Inhibitor

A. Introduction

Methods of the invention are useful for analyzing DNA from stool to detect colorectal cancer. If colorectal cancer is diagnosed early, it may be treated effectively by surgical removal of the cancerous tissue. Colorectal cancers originate in the colorectal epithelium, and typically 10 are not extensively vascularized (and therefore not invasive) during the early stages of development. The transition to a highly vascularized, invasive and ultimately metastatic cancer which spreads throughout the body commonly takes ten years or longer. If the cancer is detected prior to invasion, surgical removal of the cancerous tissue is an effective cure. However, 15 colorectal cancer is often detected only upon manifestation of clinical symptoms, such as pain and black tarry stool. Generally, such symptoms are present only when the disease is well established, often after metastasis has occurred, and the prognosis for the patient is poor, even after surgical resection of the cancerous tissue. Early detection of colorectal cancer, therefore, is important because early detection may significantly reduce patient morbidity.

B. Experiments

20 The following experiments demonstrate that EDTA, an inhibitor of DNase, increases the yield of high-integrity DNA from a stool sample with a concomitant increase in the amount of amplifiable DNA. In these experiments, three aliquots of stool (5 g each) were homogenized in buffer (0.5 M Tris, 10 mM NaCl, EDTA). The buffer to stool ratio was 7:1; thus, 35 ml of buffer

was used for each 5 g of stool. The buffer contained either 0 mM EDTA, 16 mM EDTA, or 96 mM EDTA. Each of the three aliquots was then diluted with additional buffer (not containing EDTA) to a final buffer to stool ratio of 20:1. Each aliquot was then centrifuged, and the supernatant, which carried the active DNA degrading fraction, was removed to a clean tube.

5 Then, a DNA mixture of 2  $\mu$ g *E. coli* DNA and 100 ng human genomic DNA was added to each tube. Each tube was incubated for 75 minutes at 37°C. Then, 42  $\mu$ l of Proteinase K and 250  $\mu$ l of 10% SDS (sodium dodecyl sulfate) were added to each tube followed by an overnight incubation at 37°C. After the overnight incubation, the DNA in each sample was prepared by standard techniques. *See, e.g., SHORT PROTOCOLS IN MOLECULAR BIOLOGY §§ 2.1-2.4 (Ausubel et al., 3d ed., 1995).* Generally, a phenol extraction, a phenol/chloroform extraction, and a phenol extraction were performed prior to isolating the DNA. Then, the isolated DNA was placed into a standard Tris buffer.

Three experiments were conducted on the isolated DNA. The first experiment demonstrated that the DNA degrading activity present in homogenized stool supernatant is inhibited by optimal amounts of EDTA, increasing the amount of high-integrity DNA. In this experiment, DNA was isolated from homogenized stool supernatant which was taken from aliquots of stool homogenized in buffer having 0 mM, 16 mM, or 96 mM EDTA. Total nucleic acid was run on a separation gel. Results are shown in Figure 2, where arrows identify the location of the smear (or lack thereof) containing the DNA of interest.

20 Lanes 4, 5, and 6 represent samples of DNA added to homogenized stool supernatant, obtained from stool homogenized in buffers containing 0 mM, 16 mM, or 96 mM EDTA, respectively, that was subsequently isolated. Note that each lane shows a high molecular weight band which represents endogenous DNA from the stool sample and a smear from the exogenous

DNA. The intensity of the band and smear in the photograph (which correlates with the amount of DNA in the band, a greater intensity corresponding to a greater amount of DNA) increased as the concentration of EDTA in the original buffer increased. Lanes 7-9 and 10-12 are replicates of lanes 4-6. The increasing intensity of the bands and the smears as EDTA concentration increased indicated that DNA integrity was preserved as the concentration of EDTA in the buffer increased. Thus, the DNA degrading activity of the homogenized stool supernatant was inhibited by the EDTA in a roughly dose-dependent manner.

Lanes 1, 2 and 3 and lanes 13, 14 and 15 were control samples containing 2  $\mu$ g *E. coli* DNA and 100 ng exogenous human DNA in buffer made with 16 mM EDTA. As expected, each lane showed a smear representative of the added DNA.

In a second experiment, the isolated DNA as described above was captured and amplified. The results of this experiment demonstrated that EDTA not only inhibits the DNA degrading activity present in stool supernatant but also increases the amount of amplifiable DNA. In this experiment, after preparing the DNA as described above, a standard hybrid capture was performed using Kras-specific capture probes to capture Kras DNA. The Kras DNA then was PCR amplified. Figure 3 shows the effect of EDTA on the preservation of Kras DNA. The location of the band (or lack thereof) representing Kras DNA is identified with an arrow.

Lanes 4, 5, and 6 represent Kras DNA that was amplified from template DNA that was added to homogenized stool supernatant obtained from stool homogenized in buffer containing 0 mM, 16 mM, or 96 mM EDTA, respectively. Note that the Kras band in lane 4 was nearly absent, while the Kras band grew in intensity (representing an increase in the amount of Kras DNA actually present) in lanes 5 and 6 as the concentration of EDTA in the buffer increased. Lanes 7-9 are replicates of lanes 4-6 and show a similar increase in band intensity (an increase in

the amount of DNA present) as the concentration of EDTA in the original buffer increased.

Thus, more Kras DNA was amplified, resulting in a more robust signal at higher concentrations of EDTA.

Additionally, in the population, levels of DNA which can be amplified from stool vary across individuals. These individuals have been characterized in groups from A to F, with A being the highest level of DNA and F being an undetectable level of DNA. The high levels of DNA in group A are due to low DNA degradation activity in their stool ("high-integrity stool"). Adding EDTA to the buffer in which a stool aliquot from a Group A individual is homogenized would not be expected to produce a large effect because Group A stool has little DNA degrading activity. In fact, when Kras DNA was amplified from template DNA that was added to homogenized stool supernatant obtained from Group A stool homogenized in buffer containing 0 mM, 16 mM, or 96 mM EDTA, little difference in the amount of amplified Kras DNA was observed (lanes 10, 11, and 12, respectively). Only a slight increase in Kras band intensity can be seen between 0 mM EDTA and 16 mM or 96 mM EDTA, representing only a slight increase in the amount of Kras DNA at inhibitory concentrations of EDTA.

Lanes 1-3 as well as lanes 13-15 represent samples of amplified Kras DNA that was not exposed to homogenized stool supernatant. As expected, those control lanes show a band of equal intensity across lanes representing Kras DNA. Lanes 16 and 17 are negative controls and, as expected, show no band representing Kras, indicating that any observed Kras DNA is due to captured DNA and not to contamination. Lane 18 is a negative control, and, as expected, has no band representing the Kras gene, indicating that the PCR products are from the sample and not from contamination. Lanes 19-21 are positive controls where 50 pg, 100 pg, or 200 pg of human

DNA is amplified in a background of *E. coli* DNA, indicating that human DNA can be amplified in an *E. coli* background in this model system. Finally, lane 22 is a molecular weight marker.

In a third experiment, the same protocol was used as in the second experiment except human genomic DNA was added into each sample after capture. Kras DNA was again amplified by PCR. Thus, an excess of template DNA was available for PCR. This experiment demonstrated that the varying levels of PCR amplification in the second experiment were not due to EDTA interfering with or enhancing normal PCR but were due to varying levels of template DNA available to be amplified resulting from various levels of DNA degradation inhibition by EDTA. Figure 4 shows the results of this experiment. The location of the band (or lack thereof) representing Kras is identified with an arrow. Lanes 1-6 and 8-13 correspond with lanes 1-12 in the second experiment (*i.e.*, lanes 1-3 were controls, lanes 4-6 and 8-10 were excess Kras DNA amplified in samples exposed to homogenized stool supernatant from stool homogenized in 0 mM, 16 mM, or 96 mM EDTA, and lanes 11-13 were excess Kras DNA amplified in samples exposed to homogenized stool supernatant from high-integrity stool homogenized in 0 mM, 16 mM, or 96 mM EDTA). As expected, lanes 1-6 and 8-13 show a PCR product of roughly equal intensity because an excess of template DNA is available. The EDTA does not interfere with or enhance normal PCR. This result indicates that the varying levels of PCR amplification in the 0 mM, 16 mM, or 96 mM EDTA samples in the second experiment was due to varying levels of template DNA and not to inhibitor. Lane 14 shows a sample of human genomic DNA. Lanes 15-18 are the same controls as lanes 17 and 19-21 in the second experiment.

From the experimental data described above, the amount of EDTA required to inhibit DNase was calculated. The concentration of EDTA in the various buffers used in the three

- 20 -

experiments was normalized as grams of EDTA per gram of stool. Generally, the concentration of EDTA was multiplied by the molecular weight of EDTA and by the volume of buffer in which the stool was homogenized. The product was divided by the amount of stool that was homogenized. For example, the following equations were used to normalize EDTA

5 concentration.

For 16 mM EDTA:

$$(0.016 \text{ EDTA M/L} \times 372.2 \text{ g/M} \times 0.035\text{L}) \div 5\text{g} = 0.042 \text{ g EDTA per gram of stool}$$

For 96 mM EDTA:

$$(0.096 \text{ EDTA M/L} \times 372.2 \text{ g/M} \times 0.035\text{L}) \div 5\text{g} = 0.250 \text{ g EDTA per gram of stool}$$

10 Thus, for any amount of stool to be homogenized, at least about 0.042 g EDTA per gram of stool should be used in the homogenization buffer in order to maximize yield of DNA. The range of EDTA which may be used is from about 0.042 g EDTA per gram of stool to about 0.782 g EDTA per gram of stool. More preferably, about 0.250 g EDTA per gram of stool to about 0.521 g EDTA per gram of stool is used. Most preferably, about 0.391 g EDTA per gram of

15 stool is used.

These calculations indicate that at commonly used buffer volumes and stool amounts, the amount of EDTA present in the homogenized sample is a more important factor than the final concentration of EDTA in the homogenized sample. However, as one skilled in the art realizes, at some point, although the amount of EDTA will remain the same in a given volume, the 20 volume may become so large that the effect of EDTA on DNA integrity is diluted. When examining a stool sample within commonly used parameters, this dilution effect is not seen. However, in alternative embodiments, the concentration of EDTA is a relevant factor. In these embodiments, from about 16 mM EDTA to about 300 mM EDTA is useful. More preferably,

from about 100 mM EDTA to about 200 mM EDTA is useful. Most preferably, about 150 mM EDTA is useful.

**What is claimed is:**

1. 1. A method for preserving the integrity of DNA in a stool sample, the method comprising:
  2. exposing the stool sample to an amount of an inhibitor of DNA degradation sufficient to
  3. produce a critical number of molecules of analyzable DNA.
1. 2. The method of claim 1 wherein the inhibitor comprises an ion chelator.
1. 3. The method of claim 2 wherein the ion chelator comprises a divalent ion chelator.
1. 4. The method of claim 3 wherein the divalent ion chelator is selected from the group
  2. consisting of EDTA and EGTA.
1. 5. The method of claim 1 wherein the step of exposing the stool sample to an amount of an inhibitor of DNA degradation comprises providing EDTA in a range from 0.042 g EDTA per gram of stool to 0.782 g EDTA per gram of stool.
1. 6. The method of claim 1 wherein the inhibitor of DNA degradation comprises an enzyme inhibitor.
1. 7. The method of claim 1 further comprising a step of extracting a target DNA.
1. 8. The method of claim 7 wherein the step of extracting comprises a phenol-chloroform extraction.
1. 9. A method for preserving the integrity of DNA in a sample containing exfoliated cells, the method comprising:
  3. exposing the sample to a minimum amount of an inhibitor of DNA degradation sufficient
  4. to produce a critical number of molecules of analyzable DNA.
1. 10. The method of claim 9 wherein the sample is obtained from stool.
1. 11. The method of claim 9 wherein the inhibitor comprises an ion chelator.
1. 12. The method of claim 11 wherein the ion chelator comprises a divalent ion chelator.

- 1 13. The method of claim 12 wherein the divalent ion chelator is selected from the group
- 2 consisting of EDTA and EGTA.
- 1 14. The method of claim 10 wherein the step of exposing the sample to a minimum amount
- 2 of an inhibitor of DNA degradation comprises providing EDTA in a range from 0.042 g EDTA
- 3 per gram of stool to 0.782 g EDTA per gram of stool.
- 1 15. The method of claim 9 further comprising extracting a target DNA.
- 1 16. The method of claim 9 wherein the step of extracting comprises a phenol-chloroform
- 2 extraction.

1/5

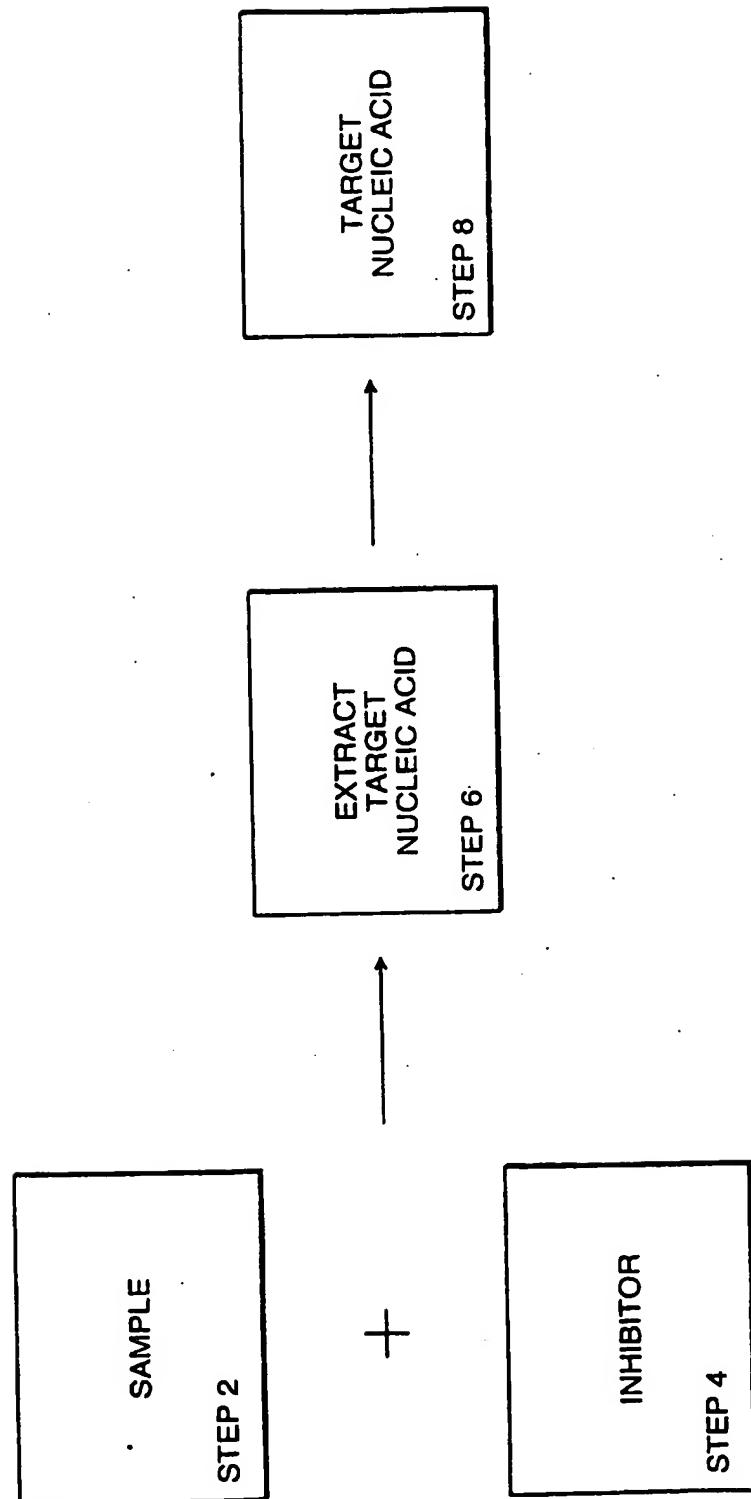


FIG. 1

2/5

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

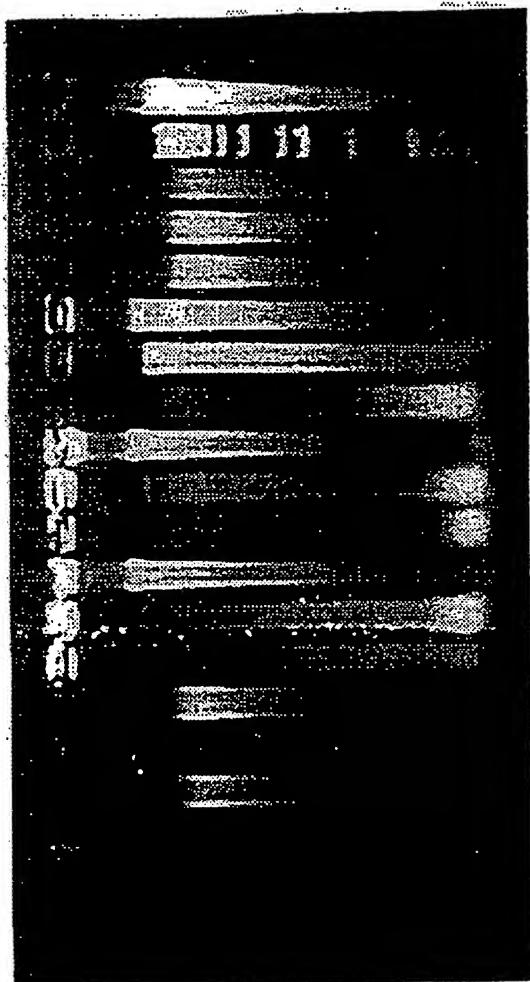


FIG. 2



3/5



FIG. 3



SUBSTITUTE SHEET (RULE 26)

4/5

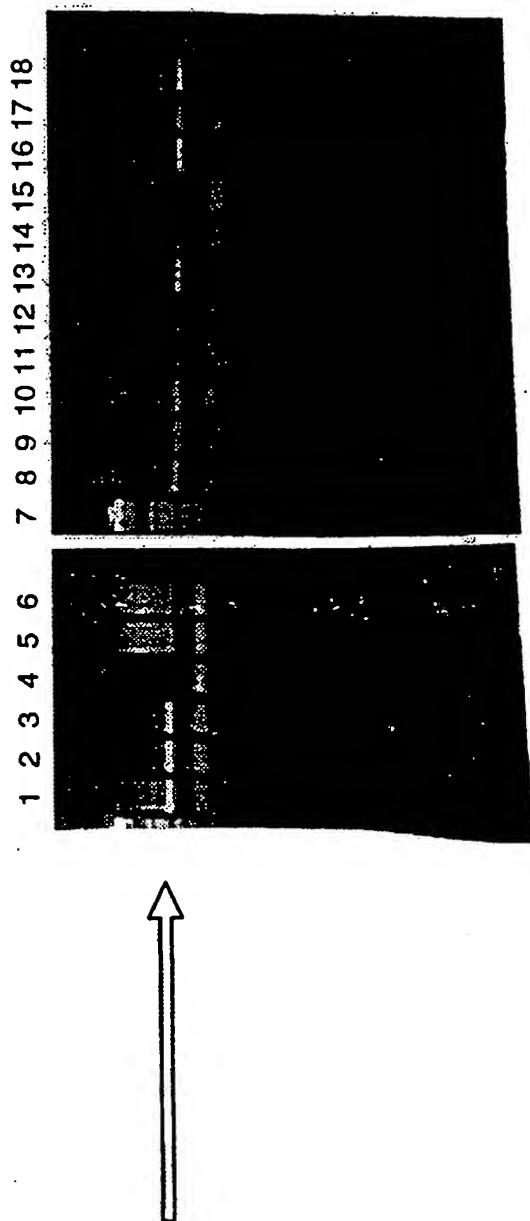


FIG. 4

SUBSTITUTE SHEET (RULE 26)

5/5

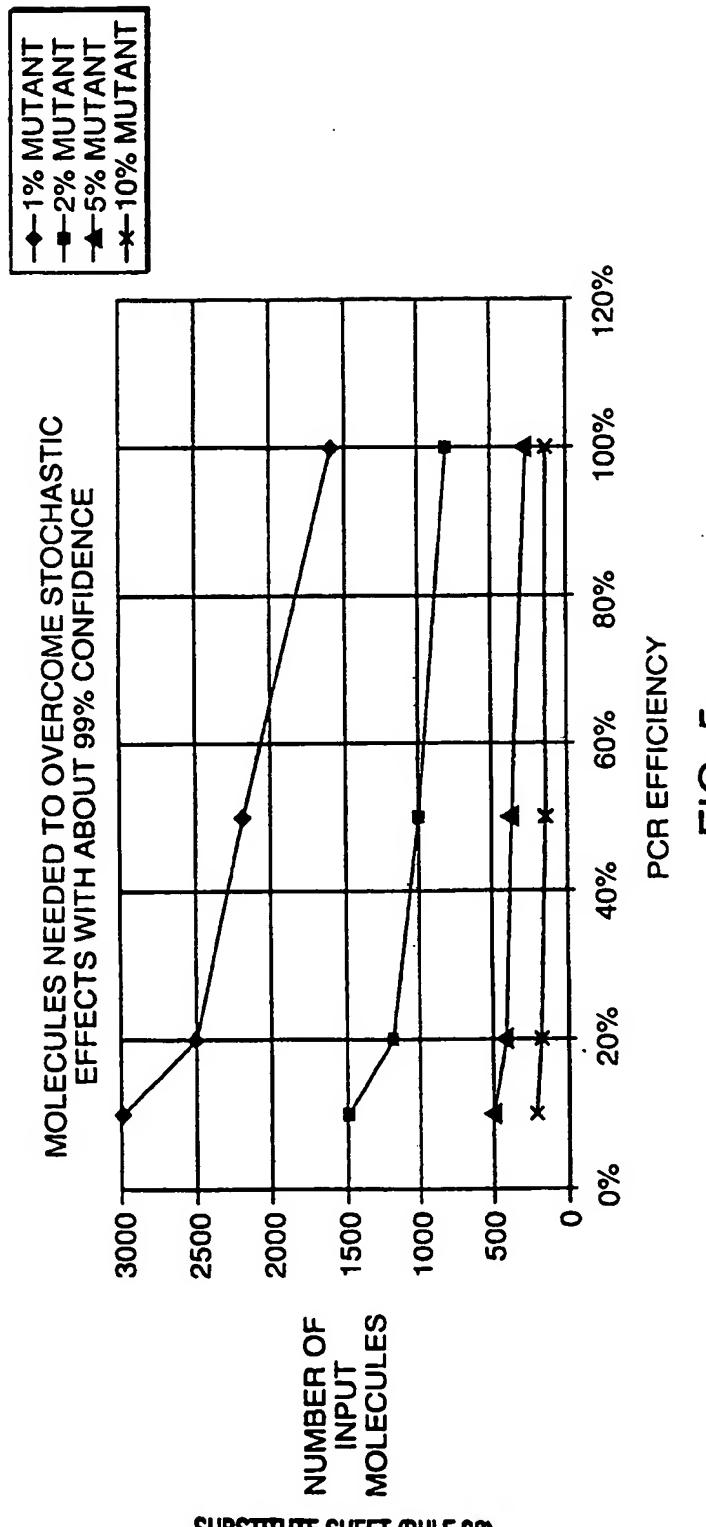


FIG. 5

SUBSTITUTE SHEET (RULE 26)

# INTERNATIONAL SEARCH REPORT

In national Application No  
PCT/US 00/01759

**A. CLASSIFICATION OF SUBJECT MATTER**  
IPC 7 C1201/68

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)  
IPC 7 C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data, PAJ, MEDLINE, CHEM ABS Data, BIOSIS, EMBASE

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 93 20235 A (UNIV JOHNS HOPKINS MED) 14 October 1993 (1993-10-14) the whole document -----	1-16
X	WO 98 58081 A (EXACT LAB INC) 23 December 1998 (1998-12-23) the whole document -----	1-16
X	DE 195 30 132 A (MAX PLANCK GESELLSCHAFT) 20 February 1997 (1997-02-20) the whole document -----	1-16 -/-

Further documents are listed in the continuation of box C.

Patent family members are listed in annex.

\* Special categories of cited documents :

- \*A\* document defining the general state of the art which is not considered to be of particular relevance
- \*E\* earlier document but published on or after the international filing date
- \*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- \*O\* document referring to an oral disclosure, use, exhibition or other means
- \*P\* document published prior to the international filing date but later than the priority date claimed

\*T\* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

\*X\* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

\*Y\* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

\*A\* document member of the same patent family

Date of the actual completion of the international search

Date of mailing of the international search report

28 June 2000

04/07/2000

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentstaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl  
Fax. (+31-70) 340-3016

Authorized officer

Hagenmaier, S

## INTERNATIONAL SEARCH REPORT

Int'l. Application No
PCT/US 00/01759

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>CALDAS C ET AL: "DETECTION OF K-RAS MUTATIONS IN THE STOOL OF PATIENTS WITH PANCREATIC ADENOCARCINOMA AND PANCREATIC DUCTAL HYPERPLASIA"            CANCER RESEARCH, US, AMERICAN ASSOCIATION FOR CANCER RESEARCH, BALTIMORE, MD,            vol. 54, 1 July 1994 (1994-07-01), pages 3568-3573, XP002050203            ISSN: 0008-5472            the whole document</p> <p>---</p>	1-16
X	<p>SIDRANSKY ET AL.: "IDENTIFICATION OF RAS ONCOGENE MUTATIONS IN THE STOOL OF PATIENTS WITH CURABLE COLORECTAL TUMORS"            SCIENCE,            vol. 256, 1992, pages 102-105, XP000605488            the whole document</p> <p>---</p>	1-16

# INTERNATIONAL SEARCH REPORT

Information on patent family members

Int. Application No  
PCT/US 00/01759

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
WO 9320235 A	14-10-1993	CA 2132874 A		14-10-1993
		EP 0672181 A		20-09-1995
		JP 8504081 T		07-05-1996
		US 5910407 A		08-06-1999
WO 9858081 A	23-12-1998	AU 7816898 A		04-01-1999
DE 19530132 A	20-02-1997	AU 712331 B		04-11-1999
		AU 6821696 A		12-03-1997
		CA 2228769 A		27-02-1997
		WO 9707239 A		27-02-1997
		EP 0851937 A		08-07-1998
		JP 11511020 T		28-09-1999